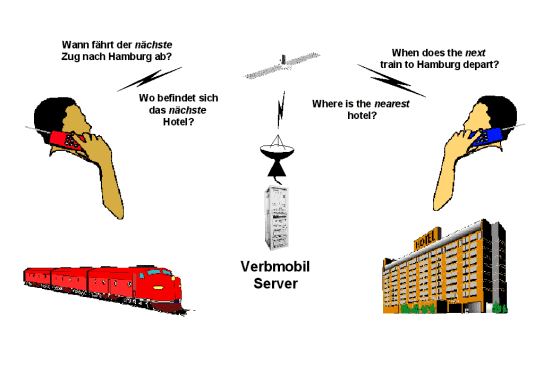
Arwa Al-Hajji

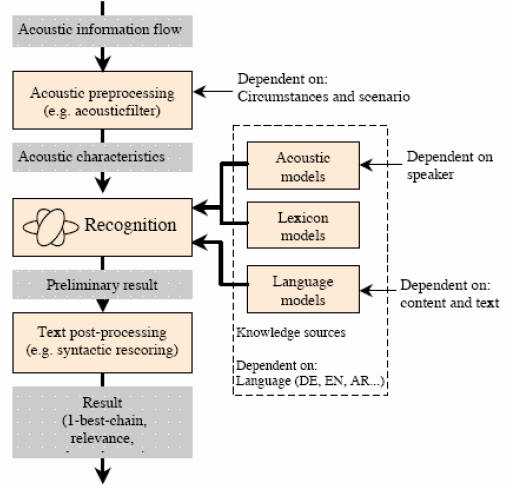
120068038



To be honest, I was so surprised when I watched the video you presented in the class. It is a common sense that interpretation is a highly specialized domain that not any one is capable to do so. How can this machine understand and produce speeches so fast while I struggle in the consecutive class to interpret 2-minute speech? It is really an amazing technology, speech recognition and voice recognition; how they differ? How do they do? This is what I tried (as well as I could) to answer thorough this paper.

***Speech recognition*** can be defines in general as a system that converts speech to words. It is an umbrella for many processes of which speech recognition is a part, but In the domain of MT, speech recognition is widely used for speech to speech that process natural language, analyze it, translate the speech to the target language and finally utters the translation.

This is a diagram that shows the process of speech recognition

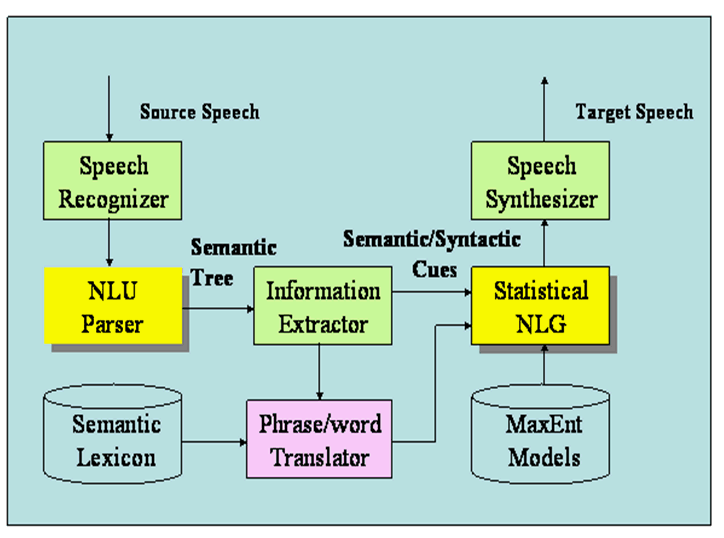


[[1]](#footnote-2)

The goal of this method is to be used to enable speech translation in real time, via recognizing the spoken words and then translating them to another language. The basic step here is speech recognition, which means recognizing the words by natural language and then to be processed by a machine that can be computer, mobile phone or another device. The process by which recognizing the speech is occurred is not that simple; this requires having the ability to recognize the words , the structures and to speak in a specific domain for which the device has a system(the sub-language Mrs. Nadia ☺)

I am providing this diagram for a specific program called IBM MASTOR System as just an example to show how this kind of technology works:

[[2]](#footnote-3)



Whereas speech recognition is used by any speaker, ***voice recognition*** is a different matter; it is training the recognition system to a particular speaker's voice, for a specific type of vocal cords product. It is more effective by training, the more you "train" your system, the more effective the recognition you will have.

I am providing the *use of speech recognition and voice recognition* here:

Medical domain:

It helps a lot in communicating with deaf people. A speaker may speak while the speech recognition system recognizes the words, and then they may be turned to another form, such as speech to text, so deaf people can understand what the speaker says. I found also that "Speech recognition can be implemented in front-end or back-end of the medical documentation process." And it also "Many Electronic Medical Records (EMR) applications can be more effective and may be performed more easily when deployed in conjunction with a speech-recognition engine. Searches, queries, and form filling may all be faster to perform by voice than by using a keyboard"

Military domain:

"Speech recognizers have been operated successfully in fighter aircraft with applications including: setting radio frequencies, commanding an autopilot system, setting steer-point coordinates and weapons release parameters, and controlling flight displays" However, according to Wikipedia website, it is not applied successfully yet, because of the many limitation that are still there.

Helicopters:

It is used in helicopters though the problem is greater in this domain because of the nose that mostly accompanies the speaker in the helicopter. However, it is used in some countries such as France, Canada and USA mostly in "control of communication radios; setting of navigation systems; and control of an automated target handover system."

Battle Management:

Commanders usually need to access large databases of information, so there is an increased machine-human interaction, and that has to be done as fast as possible. All these reasons made the use of speech recognition un urgent need; however, it is still more effective in recognizing single words, but is less effective in continues speech, but this problem is largely under search .

People with disabilities:

This technology is widely used by people who have disabilities that prevent them from using their hands. It benefits people who can't use the keyboard for example, to type voice commands. Moreover, it may help deaf people in turning voice mail to texts, or in helping people who have learning difficulties.

Telephone and other devices:

This is the area we are most familiar with. It is recording voice commands like in voice dialing. Another area is word processor, such as that we find in Microsoft Word Processor, which recognizes words to type them (speech-to-text)

***Limitation of speech recognition (and voice recognition):***

* It depends mostly on the accuracy of the speech produced by a speaker; errors such as the grammatical ones may make problems.
* This technology is more effective in recognizing individual words; continuous speech may create some difficulties.
* Noises that may be found in the environment may also create barriers to accurate recognition
* It needs controlled language, the system of the speech recognition usually has specific domain with special vocabularies such as travel and business.
* More problems can be found with the unconventional use of language such as idioms clichés.
* In other domains that need careful, such as military commanders and pilots, low- accurate speech recognition may make people avoid using speech recognition because of error possibilities.
* Different array of individual speech patterns and accents
* In the case of voice recognition, the effectiveness of the system depends on the time you spend in training it.

**Example of text-to-speech online program:**

ABC2MP3 website: <http://www.abc2mp3.com/>

Some features of this program (according to the website):

* You can choose between typing the text you want to listen to, or uploading
  + Upload MS Word
  + Upload Adobe PDF
  + Upload HTML File
  + Write Text Content
  + RSS reader
* Then, you can choose between the various voices of the speaker you would like
* You can save the file as MP3 that you download to your computer
* You can choose to share it with other people

**Example of speech to text program:**

Wave to Text

http://www.topshareware.com/Wave-To-Text-download-4950.htm

Features: (Quoted from the website)

Wave to Text is an English speech recognition-based dictation pad with a WAV to text converter. The dictation pad lets you convert your voice to text in real-time, while the wizard enables you to convert your Audio WAV files (speech recorded) offline. This speech utility is probably the most high speed way to convert speech to text accurately, no computer conscious dictation and you also don't miss a word.

**Arabic enabled speech recognition software**

I chose Sakhr software; first, because I honestly could not find another, and also because of the features that I found pioneered in the field. Some of the features are :

"The ASR engine enables IVR systems to understand people's natural Arabic speech through spoken words in different Arabic accents. Developed for telephony speech applications, the recognition engine provides speech recognition and understanding for human-like telephony services. "Sakhr's ASR 4.0 allows the developers to easily speech-enable their telephony applications for the Arabic market and with its high accuracy enhances the quality of service while increasing the customers' satisfaction"[[3]](#footnote-4)

***An area of speech to speech system:***

This is a speech recognition system that is used to translate bidirectional between two languages. It is called Multilingual Automatic Speech-to-Speech Translator (MASTOR) system. It is a "highly robust system for limited domains. MASTOR currently has bidirectional English-Mandarin translation capabilities on unconstrained free-form natural speech input with a large vocabulary (over 30,000 words for each direction) in multiple domains, including

* travel
* emergency medical diagnosis
* defense-oriented force protection and security

MASTOR runs in real-time on a laptop, and has also been ported to a handheld PDA, with minimal performance degradation."[[4]](#footnote-5)



1. http://aramedia.com/plainspeech.htm [↑](#footnote-ref-2)
2. http://domino.watson.ibm.com/comm/research.nsf/pages/r.uit.innovation.html [↑](#footnote-ref-3)
3. http://aramedia.com/asr.htm [↑](#footnote-ref-4)
4. http://domino.watson.ibm.com/comm/research.nsf/pages/r.uit.innovation.html [↑](#footnote-ref-5)